DDMA FY ’06 V&V Progress and Results
Data Driven Modeling and Analysis Team

The following DDMA work was wholly or in part funded by V&V Basic Research FY ’06 funding. The focus is on work done in the later part of the year. Work done earlier in the year was discussed in our Q1 FY ’06 report (DDMA Q1 Research Summary. doc , enclosed). Note that in FY ’06 F&F funding for radiography and metrics was combined.

- TITANS support for 4 DDMA team members
- Image and Data Analyzer (IDA) infrastructure development
- Radiographic Reconstruction
  - Deblurring
  - P < 1 Regularization
  - Compressed Sampling
  - Algorithm Improvements
- Image Pre-processing
  - Lobe finding techniques
  - Segmentation techniques
  - Boundary Parameterization Techniques
    - Trigpoly development
    - Bill’s Boundary Finder
- Image Metrics
  - Investigations of Metric Stability
  - Exploration of a broad range of metrics
  - In depth investigation of 3 classes of measures/metrics:
    - Warping methods
    - Level Set Methods
    - Whole Image Measures
- Evaluating Measures and Metrics
  - Visualization
  - Correlation
  - Least Squares Methods

The team this year was comprised of 7 LANL Staff members, 4 LANL post-docs, 5 students and more than 13 funded and unfunded collaborators from institutions across the country. The work described includes contribution from all of these people. The team members are listed in team.pdf.

1 TITANS

The motivation for the work with metrics and radiography is to improve our understanding of the physics models in the ASC codes. In FY ’06 DDMA supported 4 team members to attend TITANS, part of this money came from V&V Basic research. Our objective in attending TITANS is to better integrate our work with the objectives of the designers. Matt Sottile is a full participant in TITANS, Bryan Rasmussen, Katharine Chartrand and Rick Chartrand audit.

2 Image and Data Analyzer (IDA)

IDA is a repository for the algorithms produced by the team. It facilitates transfer of the capabilities we develop within the team and to our clients. It is also an environment for data analysis. Key elements of IDA include:

- Scripting language
- Sourceforge distribution
- Bug-tracking
- Coding standards
- Database for results
- GUI interface

The IDA infrastructure development this year touched on all of the above elements. We completed, improved and debugged:

a. a database structure to manage intermediate and final results
b. a branching data processing pipeline to support analysis sequences that are more complex than basic linear structures.
c. a scripting method for defining and recording the pipeline for a particular data set (the .ida configuration file)
d. programming standards/formats to allow new algorithms to be integrated into IDA easily and seamlessly
e. the integration of IDA into the SourceForge project management structure including bug tracking
f. documentation
g. a GUI interface

For access to IDA on Sourceforge, contact Matt Sottile (matt@lanl.gov).

2.1 Supporting Documentation

The following documents and files provide additional information about this year's IDA development and are included with this report.

IDATopTen.doc
Every now and then we ask ourselves why we are developing IDA. This document lays out our objectives for the project .

IDA Path Forward.pdf
This is an in-depth discussion of where IDA development stands and key areas for improvement, ranked by importance and difficulty. This documentation is the basis for our collaboration with the University of Oregon. They have agreed to help us with IDA development in exchange for the use of the code for their own projects. This document covers near-term focus areas where IDA requires improvement on existing code, along with long-term research topics that will push

IDA ahead and address features that are not currently present in tools for image and data analysis.

Ida-0.8.2
A directory containing most recent IDA release. The documents sub-directory contains the current IDA documentation. The stable_code directory contains the individual algorithms. The algorithms are documented in the headers of the code.

3 Radiographic Reconstruction

DDMA's TV Abel radiographic reconstruction algorithms are state-of-the-art and fast and available upon request to LANL staff. Contact Rick Chartrand (rickc@lanl.gov).

3.1 Better Reconstructions

3.1.1 Shape-based reconstruction

Many LANL simulations are judged by their ability to predict shapes. However, the experimental reconstructions to which they are compared are not produced in a shape-faithful manner. Both reconstruction methods in use, the BIE and DDMA's TVAbel, penalize certain aspects of image shapes. In the case of the BIE, the penalization is implicit and difficult to describe or rectify. In the case of TVAbel, it is simply that total variation penalizes the perimeter of shapes.

Our research this year put us in a position to improve the TV Abel algorithms and generate experimental reconstructions that preserve shapes accurately. A modification of total variation will penalize only higher-dimensional (fractal) shape characteristics, and not perimeter, while still suppressing noise. This can make the V&V process meaningful in a manner it has not been before. This comes at a cost of presenting a more challenging optimization problem, but the issues are manageable, making shape-based reconstruction a near-term capability.

3.1.2 Accurate noise modeling

Standard regularization methods in the literature are often best suited for additive, Gaussian noise. Our team completed work this year on adapting existing models to better handle the types of noise more commonly found in experiments. The two primary examples are signal-dependent (Poisson) noise and impulse noise. Ongoing work is needed to fully implement these new models into our computational framework and apply them to relevant problems.

One example of this type of work is our research on a class of duality-based algorithms for total-variation regularization with general data fidelity. Total-variation regularization involves a non-differentiable penalty term, which makes implementations numerically difficult. Conventional approaches use an approximation to TV, which degrades some of the characteristics of TV that make it useful (such as preserving jump discontinuities).

A duality-based algorithm of A. Chambolle allows the exact-TV reconstruction to be computed. However, this has only been done for a mean-squared-error data fidelity term. This is only the best choice if the noise is additive Gaussian. Previous work identified a data-fidelity term that gives better results for Poisson noise. Furthermore, the Chambolle algorithm does not allow the incorporation of an operator for the purpose of inversion, such as deblurring or Abel inversion.

We have remedied this by generalizing the Chambolle algorithm for a very general class of data fidelity terms. We have also proved that the algorithm converges for a sufficiently small timestep. Our convexity-based approach will allow for further generalization.

Many LANL data exhibit Poisson noise, making MSE data fidelity terms suboptimal. Complicated, discontinuous structures are best recaptured with true TV regularization, rather than the approximation of conventional approaches.

3.1.3 Reducing artifacts of the algorithms in reconstructions

Graduated Adaptive Image Denoising is an improvement to the TV Abel reconstruction method that reduces artifacts caused by the algorithms. This method introduces variants of the variational image denoising method proposed by Blomgren, Mulet, Chan, and Wong, which interpolate between total-variation denoising and isotropic diffusion denoising. We have investigated how parameter choices affect results and allow tuning between TV denoising and isotropic diffusion for respecting texture on one spatial scale while denoising features assumed to be noise on finer spatial scales. Furthermore, we proved existence and (where appropriate) uniqueness of minimizers. We considered both L2 and L1 data fidelity terms.

3.2 Making the Most of the Data We Have

3.2.1 Using dynamics information

The use of dynamics information to enhance tomography has been studied in simple contexts by our team. We can reconstruct two-dimensional objects from one-dimensional multiple-time projections given the parameterized dynamics that connect the objects at the various points in time. This work was performed on several illustrative basic problems and has been accepted for publication. This work needs to be extended to include complex 2d object descriptions, realistic radiographic projections, and more detailed physical models of the dynamics.

3.2.2 Severely underdetermined inversions of 3D information

This year a visit by Emmanual Candes precipitated extensive research within the DDMA team in an area known as “compressed sensing.” Compressed sensing provides the capability to obtain high-quality reconstructions from far less data than previously thought possible. Our research in this area has reduced still further the amount of data necessary. A few of the many potential applications are as follows:

- 3-D reconstructions from few (even two) radiographs: The methods of compressed sensing provide a way to dispense with the cylindrical symmetry assumption required by Abel inversion, while still needing far fewer projections than traditional 3-D inversion methods. Combined with an assumption that an object is nearly symmetric, these methods may enable 3-D reconstructions from dual-angle radiographs. By combining these techniques with the inclusion of dynamics information, we can make optimal use of the dual-angle and multiple-time radiographs envisioned from DARHT.
- 3-D corrections to single, asymmetric radiographs: The compressed-sensing approach can allow the symmetry assumptions required by Abel inversion to be relaxed. This will permit better reconstructions from single radiographs exhibiting 3-D effects. Similarly, these methods provide the flexibility to ignore artifacts or portions of bad data, without requiring sometimes clumsy preprocessing.

3.3 Deblurring

Deblurring is the process of inverting the convolution that creates image blur. Inverting the convolution operator naively can amplify noise to the point of uselessness, so we regularize the process by constraining the result to belong to a particular class of (noiseless) images. This year we investigated the application of two regularizations to classified problems for Alan Mathews:

3.3.1 Total variation (TV)

Requiring the image to have a finite TV eliminates noise, but, crucially, does not smear edges. It does, however, eliminate small-scale features, particular edges with long length but containing little area (such as tendrils).

3.3.2 Poisson singular integral

This uses a two-parameter family of Besov-type spaces that are larger than the space of functions of total-variation. This weakening of the regularization will permit the recovery of small-scale features not retained by TV.

3.4 Faster Reconstructions, Better Analysis

The DDMA Image and Data Analyzer (IDA) is an environment for data analysis that automates many processes needed to perform accurate radiographic inversions. The algorithms within the package incorporate state-of-the-art techniques, in some cases before they appear in the literature. These tools are available to the lab community on Sourceforge, either as a part of IDA or separately for inclusion in other contexts.

While highly effective, TV regularization is very computationally expensive. This expense grows with the complexity of the inverse problem being regularized, significantly restricting the size of problem that can be solved in cases such as fan-beam and tilted-beam radiography. Efficient algorithms are critical in making this approach feasible for such data sets.

Our current algorithms work on the order of seconds to hours, depending on the problem. This year we improved the speed of some of our algorithms to allow us to perform parameter studies more efficiently. This is a step toward improving our ability to repeat inversions over many perturbations of parameter choices to quantify the uncertainty associated with the algorithms. We are well poised to develop faster methods; one of our post-docs, Paul Rodriguez had the world's fastest FFT implementation at one time.

3.5 Supporting Documentation

The following documents provide additional information about this progress in radiography and are included with this report.

TV Primer.pdf
Background on our TV Abel regularization research written for DDMA students.

Reconstruction of sparse signals.pdf
Chartrand, Rick. "Exact reconstruction of sparse signals via non-convex minimization." – Submitted.

Nonconvex compressed sensing.pdf
Chartrand, Rick. "Nonconvex compressed sensing and error correction." – submitted.

Nonconvexreg-slides.pdf
Slides on the research in the above two papers

Schultz-2005-graduated.pdf
Peter F. Schultz, Erik M. Bollt, Rick Chartrand, Selim Esedoğlu and Kevin R. Vixie, "Graduated adaptive image denoising: local compromise between total variation and isotropic diffusion", Submitted., 2005

This paper discusses methods we have developed to reduce artifacts of the algorithm in TV Abel reconstructions

Le-2005-variational.pdf
Triet Le, Rick Chartrand and Thomas J. Asaki, "A variational approach to reconstructing images corrupted by Poisson noise ", Submitted., 2005

A denoising technique we applied to problems posed to us by Alan Mathews.

Carasso.pdf
A deblurring technique we investigated this year that does an exceptional job of recovering detail.

4 Image Pre-processing for Measures and Metrics

Many methods for quantifying a characteristic of an image as a measure require pre-processing the image to extract features and shapes of interest. This is a critical but not glamorous area of work where we made a lot of progress this year. In particular, we have developed methods to examine the effect of preprocessing on the robustness, stability, and uncertainty that is induced by algorithmic decisions on the end measure or metric. We are actively pursuing methods that we have identified as rigorous, defensible ways to derive such quantitative uncertainties that result from pre-processing methods.

4.1 Lobe Finding Techniques

If one is investigating shapes that resemble butterflies, one might be interested in the number, order and characteristics of the lobes corresponding to wings and antennae. The problem here is finding a reliable technique. Once lobes have been clipped from the central body they can be analyzed with a number of very useful simple measures. We developed two approaches to clip the lobes off a shape.

The first method uses the inverse of the convex hull computed on an “inside-out” transformation of the whole shape.

In the second method, the boundary is parameterized as a trigonometric polynomial (Fourier-like description) using the IDA function trigpoly and spectral features of the shape are used for lobe clipping. The low-frequency boundary approximation (by low-pass filter) defines a boundary of either a maximal spatial frequency or fractional cumulative power. By cataloging the crossings of this curve and the original boundary, the lobes are easily discriminated. Segmentation of the shape into lobes and body is then performed. Protruding lobes of simply-connected 2-d regions can be quickly identified and segmented in this way.

4.2 Segmentation

We investigated several segmentation methods this year for stability and capabilities. In particular, we made a thorough study of k-means and the hierarchical application of this method to identify features in complex images. This technique was very successful on jets data provided by Bernie Wilde. We have also implemented a segmentation technique that has recently appeared in medical imaging literature that, unlike common methods such as k-means, can be tuned to balance spatial relationships between pixels with global intensity distributions. This method, spatial fuzzy c-means, performs very well in the presence of noise without requiring denoising before segmentation. We have extended this algorithm to use an isotropic spatial weighting window that was not present in the published method.

This method is of particular interest for use in our ongoing and future work, as it is able to preserve features without distortion more accurately than some methods based on a combined preprocessing-before-segmentation process. Example images, both physics-based and photographic, are shown in the attached wildemay.pdf progress report for the Jets project.

4.3 Boundary Parameterization

4.3.1 Trigpoly

In many image analysis tasks it is beneficial to represent a region by its boundary. A good boundary description should be as independent as possible from any pixelized representation since pixelization is entirely an artifact of a detection or discretization process rather than of inherent object properties. The routine trigpoly represents a binary image by the trigonometric polynomial coefficients needed to adequately reproduce the boundary. Low pass filtering on the coefficients produces boundary representations that are smooth relative to pixel-size features. Trigpoly is currently the workhorse of our boundary parameterization techniques.

In the course of our work this year, we identified and addressed some undesirable features in this algorithm. First, disconnected regions were not well handled. Trigpoly was able to identify multiple regions, but no global description was available. Second, trigpoly could not evaluate images with pixel-description boundaries that were not unique. This can occur when objects have a lot of pixel-size detail or when a simple boundary is corrupted by noise. These two problems have been addressed. Trigpoly can now return individual boundary descriptions or a description based on an artificially connected global region. Trigpoly also now can modify binary images to remove all ambiguities in the boundary path.

It is quite possible that other boundary parameterization methods will suffer from similar problems when a shape derived from an image contains multiple disconnected regions. For a single image, one can provide some relationship between the regions (such as artificial bridges between regions), but this relationship does not necessarily translate to other images. Furthermore, no boundary parameterization currently in use takes advantage of the dynamics in the source process that is represented in the images should they be related by time. A simple example of this would be a time series of images of a fluid droplet separating into three separate drops. Current methods may choose to connect them in a manner that varies (detrimentally to a metric) over time.

4.3.2 Boundary Approximation by Triangulation

We developed an alternative boundary approximation method that uses geometric principles to find a boundary. The details of this method are included in boundarybytriang.pdf, attached. This method will allow us to avoid spurious oscillations that may appear for some images in a shape boundary due to the trigonometric polynomial parameterization.

4.3.3 Supporting Documentation

Boundarybytriang.pdf
A discussion of Bill Allard's boundary finding technique.

Wildemay.pdf
A slide show which summarizes our investigation into segmentation techniques. Slides 11-17 show the comparison of k-means, fuzzy c-means, and spatial fuzzy c-means. Slide 18 compares a smoothing operation before k-means (left image) with the output of spatial fuzzy c-means to illustrate the potential extreme effect of preprocessing on subsequent segmentation.

5 Measures and Metrics

5.1 A hundred measures (at least)

We investigated and applied very large number of simple measures to relevant problems and then evaluated them by eye or by correlation with other relevant characteristics of a model. These include:

1. Area.
2. Perimeter.
3. Volume of rotation.
4. Surface of rotation.
5. Isoperimetric ratios.
6. n-th moments (i.e., (statistical moments) of intensity histograms
7. Uniformity and entropy of intensity histograms.
8. Betti numbers.
9. "Pointwise distance" histograms. (Histogram of distance between all perimeter points.)
10. Eccentricity.
11. Area of convex hull.
12. Compactness.
13. Sharpness.
14. Number of turns in perimeter.
15. Integral of absolute value of turns in perimeter.
16. Statistics on scalar measures taken across all level sets. (Example: Skew of Area as a function of threshold.)
17. Orientation.

18. Major/minor axis lengths.
19. Connectivity.
20. Metrics from templates.
21. Ratios and combinations of all of the above.

5.2 Metric Stability

We investigated the sensitivity of our metrics to algorithmic decisions and found in some cases our segmentation and boundary finding techniques were sensitive to parameter choices unrelated to the data. In addition, we were posed a number real problems this year where the structure we were evaluating did not have a single relevant boundary. This strongly informed our investigation of new metrics. In particular, we sought metrics and measures that do not depend on finding a single boundary, in addition to beginning an investigation into methods for quantifying these effects as discussed earlier.

5.3 Investigating new measures metrics

This year, we investigated twenty or thirty directions for measures and metrics and ultimately focused on three areas with the greatest opportunity for weapons applications: warping methods, level set methods and whole image metrics.

5.3.1 Warping

A warp is a transformation of one signal or image into another. The transformation is performed on the domain in which the data is embedded (eg: 1D sample spacings, a 2D image mesh, etc…), with a comparison against a second data set based on a domain remapping of the warped image followed by a basic norm. Once a transformation is identified that minimizes this norm, the cost of the warp is computed based on the energy or work required to perform the transformation itself. The cost of the warp can be a metric of the difference between the two data sets based on the difficulty of turning one into the other in a physically motivated way. Warping can be used to compare two sampled signals, two vector-type measures on an image or the images themselves.

Warping methods are gross metrics of the differences between two images. Like L2, they don't extract any particular characteristic in the data. However, warping can tell you the cost of differences in registration, dilation, translation and contraction of the shapes that L2 will miss. In general, warping metrics reproduce well what the mind does in measuring gross differences between shapes and signals. Warping; however, is computationally challenging and requires considerable investment in programming and testing. This year we looked closely at three types of warping: elastic warping, fluid warping and dynamic time warping.

5.3.1.1 Elastic Warping

Elastic warping produces a quantitative cost based on the elastic energy of the transformation that can be thought of as the amount of work required to turn one image into the other using vibration modes of a sheet of rubber. This method, although conceptually attractive, has proven to unusable in practice so far for two reasons. First, determining the transformation (ie, identifying the frequencies and amplitudes of the vibration modes) is a difficult optimization problem that is computationally intensive. Second, the vibration modes are a global property of the warp that makes it difficult to represent multiple small, local and spatially separated warps without resorting to a huge number of frequency components.

This second problem ties in with the first, as small local warps are often necessary in practice. Optimization routines have difficulty identifying them due to limited a-priori information about coefficient scale and relevant frequencies to focus on.

5.3.1.2 Fluid Warping

Fluid warping produces a quantitative cost based on the principles of fluid flow; one can use the differences between two images to provide the force that drives a flow. Each time step alternates between computing a new displacement field by evolving the fluid using this force, and computing the driving force for the next time step based on the difference between the warped image and the image being compared to. The computational cost of these methods is related to the method employed to solve the fluid dynamics equations for each step in evolving the flow.

Fluid warping has been used with great success in the medical imaging field for image registration purposes, and overcomes many of the difficulties that were encountered in our investigation of elastic warping. We have been able to develop this method to the point where we can begin to test it on relevant data and the results are very promising. We are also exploring how we can take advantage of the special nature of this particular fluid flow problem to relax requirements on the solver technique employed to decrease the computational overhead for performing a single warp.

5.3.1.3 Dynamic Time Warping

Dynamic time warping is an algorithm that finds an optimal match between two sequences independent of non-linear variations in the data. It originates from speech processing and word recognition in sampled audio signals. For example, if one were comparing data from speech patterns, the algorithm would detect similarities despite accelerations and decelerations in word utterances over the course of the measurement. This addresses critical problems with L2 methods including trouble with dilation or contraction of a signal and signals of different lengths. In addition, we extended the algorithm to introduce other constraints that we are interested in and differences in registration of the signal by treating them as potentially periodic (circular) should the application area require it.

5.3.1.4 Conservative Remapping

Warp methods require a remapping operation to occur during the evolution of the warp in order to have data in compatible domains for computing the norm that drives the evolution. We have found that warping methods that use basic interpolation schemes for this remapping can suffer from ill effects due to mass (or intensity) conservation violations. We have spent some time testing the CORE (Conservative REmapper) software package from T-7 as a potential remapping operator for this purpose. CORE is quite powerful and intended for remapping of arbitrary 2 and 3-dimensional meshes. For performance reasons, we are in the process of completing development of a conservative remapper that takes advantage of assumptions that can be made about the topology of meshes derived from pixelated image data to increase computational performance.

5.3.2 Level Set Methods

We investigated a large class of measures that consist generally of taking the measures discussed in the section 5.1 and evaluating them on a series of super-level sets to get a characteristic vector for an image. These vectors are then compared using the dynamic time warping technique. This metric is much more stable relative to metrics that depend on the selection of a single level set. We have a number of promising results from this investigation. An example of this method using area is shown in the Sept 2006 VV Review.ppt under the title Symmetric Rearrangement.

5.3.3 Whole Image Measures

A whole image metric is computed on the raw data after minimal preprocessing (such as denoising, Abel inversion, etc…). Less preprocessing steps leads to fewer opportunities for processing artifacts to induce error or metric uncertainty that is not related to the data itself. No pixels are isolated as relevant (shape) versus irrelevant (background) and therefore no need to make a binary decision where one is not appropriate.

The whole image measures that we investigated include extending our geomeasures techniques to 3 dimensions. We also investigated graph mean curvature. These results are also in Sept 2006 VV Review.ppt.

Whole image measures are an area that we have just begun investigating; we plan to continue work in this area in the following year.

5.3.4 Image Measures and Metrics by Characteristic Shape

This approach finds low-dimensional measure vectors on whole images that preserve relevant features and compute metrics on feature vectors. We use low-dimensional parameterized shapes to form relevant measure vectors as a function of image intensity. This method discussed furthering CharShape.ppt

5.4 Supporting Documentation

The following documents provide additional information about this progress in radiography and are included with this report.

CharShape.ppt
Discusses image measures and metrics by characteristic shape.

Metrics-path forward.pdf
Discusses the key issues with metrics based on our experience this year and outlines our path forward.

Sept 2006 V&V Review
One approach applied to Bernie Wilde's Jet's problem and what we learned.

Warps.ppt
Why warps are important, how they work, and what we plan to do this year.

6 Evaluating Measures and Metrics

We currently have a suite of at least 100 simple measures, and on the order of 10 sophisticated ones. The question is which ones inform us about the questions that matter, and ignore features that do not.

6.1 Visualization

One common and effective way to evaluate measures is to sort images by a measure or metric and compare how the method ranks the differences between images by eye. This remains the best way to "get" what a metric is doing. We have considered a number of ways to present results in this way.

6.2 Correlation

We have investigated correlate them with other known, important features of the model. Contact Bryan Rasmussen for these results.

6.3 Least Square Fit to Expert Opinion

One interesting problem is to design a combined measure that consists of a vector of relevant measures. A useful combined measure uses only metrics that matter and weights them appropriately. We considered the problem of developing a compound measure as follows:

- We want a compound measure which:
  - Is composed of distinct, definable components, and
  - Quantifies the contribution of each component.
- We want a metric that codifies expert opinion in a systematic way, and
- We need a process that evaluates which measures are relevant to the problem and which are not.

We developed a table of expert opinion on the pairwise differences between a set of images. We then evaluated the pairwise differences between the images with 200 metrics and used a least squares fit to find the combination of metrics which most closely approximated expert opinion. We identified some issues with the process, but we believe it is a promising way to both evaluate metrics and develop combined metrics. Results on an unclassified problem are included in nonnegleastsquares.pdf.

6.4 Supporting Documentation

The following documents provide additional information about this progress in radiography and are included with this report.

Nonnegleastsquares.pdf
An example of the combined metric approach applied to an open problem.